基于局部感受野扩张 D-MobileNet 模型的图像分类方法 *

王　威，邹　婷，王　新†

(长沙理工大学 计算机与通信工程学院，长沙 410114)

摘　要： 针对轻量级的深度神经网络 MobileNet 会减少分类准确率的问题，将空洞卷积核引入 MobileNet 模型中的某一卷积层中，提出一种基于局部感受野扩张的 D-MobileNet 模型。模型根据空洞卷积核所在位置的不同分为三种结构，在不增加参数数量的同时能够扩大该层卷积核的局部感受野，提高分类精度。实验在 Caltech-101 数据集、Caltech-256 数据集以及图宾根大学动物分类数据库上进行，结果表明，D-MobileNet 模型可获得比 MobileNet 更好的分类准确率，最多可以提高 2%。

关键词： 图像分类；深度神经网络；MobileNet；空洞卷积；D-MobileNet

中图分类号： TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.10.0781

Image classification method based on D-MobileNet model

Wang Wei, Zou Ting, Wang Xin†

(*School of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha* 410114, *China*)

Abstract: Aiming at the problem that lightweight deep neural network MobileNet can reduce classification accuracy, this paper proposed an D-MobileNet (dilated convolution MobileNet) model based on local receptive field expansion by introducing dilated convolution kernel into a convolution layer of MobileNet model. The models consisted of three structures according to the location of the dilated convolution kernel. Without increasing the number of parameters, it could expand the local receptive field of the layer convolution kernel and improve the classification accuracy. This paper carried out the experiments on Caltech-101 database, Caltech-256 database and Uebingen animals with attributes database. The results show that the D-MobileNet model can achieve better classification accuracy than that of MobileNet, and can improve the classification accuracy by up to 2%.

Key words: image classification; deep neural networks; MobileNet; dilated convolution; D-MobileNet

0　引言

计算机图像分类利用计算机对图像进行分析，把图像归为若干类别中的某一类别，来代替人的视觉判读，是计算机视觉领域的研究热点之一。大多数图像分类的研究主要集中在图像特征提取和分类算法上面，特征的好坏对分类非常关键，而传统的图像特征如 SIFT、HOG 等特征都是经过手工设计的，因此有时难以满足要求。卷积神经网络具有自学习、自适应和自组织能力，能利用已知类别的图像样本集的先验知识，自动提取特征，可避免传统图像分类方法中复杂的特征提取的过程，提取到的特征表达能力强，分类效率高。

深度卷积神经网络在计算机视觉领域， 如图像分类[1]、目标追踪[2]、目标检测[3]，以及图像分割[4]等方面都取得了很好的效果。如 Krizhevsky 等人在 2012 年 ImageNet 大规模视觉识别挑战分类任务使用约有 6000 万参数 8 层的 AlexNet[1] 模型取得了冠军，Simonyan 等人使用的 16 层的 VGG[5]、以 Inception 为基本结构的 GoogleNet[6]、为改善梯度消失问题引入残差结构的 ResNet[7]等也都取得了成功。但由于深度卷积神经网络模型本身是一种结构密集型和计算密集型的模型，庞大的参数数量和计算量、大量的内存访问和 CPU/GPU 资源计算导致的巨大的耗电量使得模型难以应用到硬件资源有限的便携式移动设备上。

针对上述问题的一种可行的解决方案是对深度神经网络进行压缩与加速。在尽量不影响精度的前提下，减少网络参数和计算量，减少耗电，使完整的深度神经网络能应用到一些有实时性要求和低内存的便携式设备中。Denil 等人[8]证明了深度神经网络中的参数存在大量的冗余，且这些冗余的参数对分类精度并没有很大的影响。Denton 等人[9]通过 SVD 奇异值矩阵分解找到一个合适的低秩矩阵来估计深层 CNNs 的信息参数，该方法需要较多的计算成本，也需要大量的重新训练来达到收敛。Han 等人[10]通过参数剪枝将训练好的网络中不重要的连接删除，对剩下的参数再进行训练和量化，然后对量化后的参数进行霍夫曼编码，进一步降低压缩率，该方法需手动调超参数。Hinton 等人[11]采用知识精馏的方法对网络模型进行压缩，将一个性能好但存在较多冗余的复杂网络中有用信息提取出来迁移到一个更小更简单的网络上，使简单网络与复杂网络有相近的性能。 除此之外，很多相关研究通过改进网络模型来压缩网络。 SqueezeNet[12]是以 fire module 为基础结构的网络模型、MobileNets[13]是以深度可分离卷积核（depthwise separable filters）为基本结构的网络模型、ShuffleNet[14]的基本结构是在残差结构的基础上进行改进， 引入了分组逐点卷积（group pointwise convolution）和

收稿日期： 2018-10-07；**修回日期：** 2018-11-23　**基金项目：** 国防预研项目；国家自然科学基金资助项目 （61070040）；湖南省教育厅科研项目（17C0043）

作者简介： 王威（1974-），男，教授，博士，主要研究方向为智能信息处理；邹婷（1994-），女，硕士研究生，主要研究方向为智能信息处理；王新（1976-），女（通信作者），讲师，硕士，主要研究方向为智能信息处理（wangxin@csust.edu.cn）.

通道重排（channel shuffle）操作。

轻量级网络虽然参数或计算量减少了，但是分类准确率也有相应的下降。为了减少计算量同时又兼顾分类精度，本文提出一种基于局部感受野扩张的 D-MobileNet 网络结构，将空洞卷积引入到 MobileNet 网络中，利用空洞卷积核在不增加参数的前提下可增大卷积核感受野这一优点，获取更大的局部感受野，提高 MobileNet 的分类精度。

1 卷积神经网络基本理论

1.1 卷积神经网络

作为一种自动化特征提取的机器学习模型，卷积神经网络是最早应用于各个领域的深度学习网络模型。卷积神经网络通过局部感受野和权值共享减少权值的数量，降低网络模型的复杂度，从而避免一般深度学习模型中由于参数过多而引起的过拟合问题。与此同时，卷积神经网络可将图像直接作为网络的输入，避免传统图像分类方法中复杂的特征提取过程，并且对图像平移、比例缩放等具有高度不变性。

卷积神经网络一般由卷积层、池化层和全连接层组成，如图 1 所示。图像经一层或多层卷积层和池化层进行特征提取，将最后一层卷积层输出的所有特征图转换成一维向量进行全连接，最后利用分类器进行分类。网络通过反向传播调节权重参数，并利用分类的结果与期望输出的结果之间的平方差达到最小这一目标进行优化。卷积神经网络每层的神经元按宽度、高度以及深度三维排列，其中宽度和高度指神经元尺寸的大小，而深度指输入图片的通道数或输入特征图的数量。

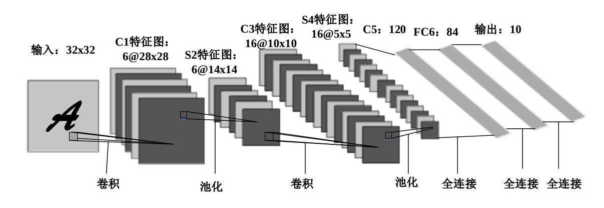


图 1 卷积神经网络基本结构示意图

Fig. 1 Basic structure of convolution neural network

卷积层通过卷积运算来提取图像中不同的特征，该层包含若干组可学习的参数(卷积核)，是整个卷积网络的核心。当前层的卷积核对输入的特征图像进行卷积运算，可提取局部特征，得到特征图，再经过激活函数进行非线性操作， 可得到非线性的特征映射图像。卷积神经网络的局部感受野和权值共享思想在卷积层中体现，网络中大部分的计算量也集中在卷积层。

卷积层之后为池化层，也叫下采样层。该层通过下采样操作，在一定大小的区域内，用一个特定的值作为输出，并通过去掉特征映射图中不重要的样本点来降低下一层输入维度，进一步减少运算量，增加网络对图像平移、旋转等变化的适应性。常见的池化操作有最大池化和平均池化。

卷积层+池化层的结构可提高网络的鲁棒性，卷积神经网络通过多层卷积层可得到更深层次特征图。随着层数增加，学到的特征也越全局化。最后学习到的全局特征映射转换成向量连接全连接层，再连接分类层。网络中大部分的参数量在全连接层。

1.2 MobileNet 模型

MobileNet 是一种流线型的架构，它使用深度可分离的卷积来构建轻量级的深度神经网络，为移动和嵌入式视觉应用提供的一种高效模型[13]。 MobileNet 的基本结构为深度可分离卷积核（Depthwise Separable Filters），如图 2 所示。

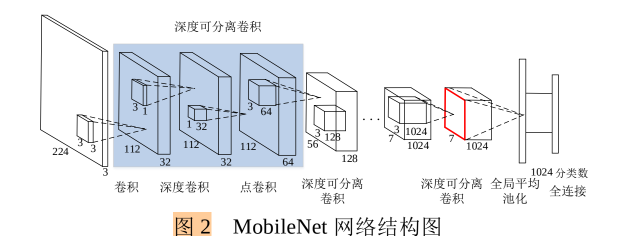
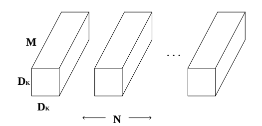


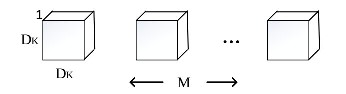
图 2 MobileNet 网络结构图

Fig. 2 Architecture of mobilrnet

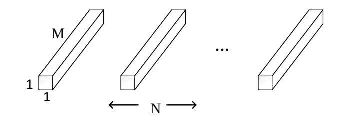
深度可分离卷积核由深度卷积核(depthwise convolution filter)和点卷积核(point convolution filter)组成，其中深度卷积核在每个输入通道上进行单个卷积，点卷积核用一个 1×1 的卷积将深度卷积的输出值进行线性组合。 这样，N 个 $D_K\times D_K\times M$ 的标准卷积核（图 3(a)）可以由 M 个 $D_K\times D_K\times 1$ 的深度卷积核（图 3(b)）和 N 个 $1\times1\times M$ 的点卷积核（图 3(c)）替代。一个标准的卷积滤波器是将输入组合成一组新的输出，而深度可分离卷积将输入分为两层， 一层用于过滤，另一层用于合并。



(a)标准卷积核(standard convolution filter)



(b)深度卷积核(depthwise convolution filter)



(c)点卷积核(point convolution filter)

图 3 标准卷积核和深度可分离卷积核

Fig. 3 Standard convolutional filters and depthwise separable filters

2 D-MobileNet 模型

标准的 MobileNet 模型一直都采用 3×3 的小尺寸卷积核。这样虽然可以减少计算量，但是在前几层特征图分辨率较高的情况下，小尺寸的卷积核的局部感受野太小，捕捉不到好的特征。若换成较大的卷积核，则又会增加参数数量和计算量。因此，可以考虑在前面几层卷积层中，用扩张率为 2 的空洞卷积代替标准卷积。这个模型称为扩张卷积 MobileNet 模型，即 D-MobileNet 模型。

2.1 空洞卷积

空洞卷积核[15]（dilated convlution）又叫做带孔卷积核，是在上采样滤波器非零值中间插入零值的一种卷积核。空洞卷积最先应用在图像分割中。图像分割需要得到与原输入图片相同尺寸的图片，而传统的深度神经网络中池化层会减少特征图的空间分辨率。为了生成有效的密集特征图，Chen 等人将全卷积神经网络去掉后面几层最大池化层，同时，为了取得相同大小的感受野而引入空洞卷积。这样既能避免池化层减少特征映射图空间分辨率，还能与池化层一样增加感受野[4]。

带孔卷积核就是通过在卷积核中非零数值中间插入零值扩大该卷积核的感受野，如图 4 所示。其中，(a)表示标准的 3×3 卷积核的感受野，(b)表示扩张率为 2 时不加填充时 3×3 卷积核的感受野为 5×5，(c)表示扩张率为 3 时不加填充时 3×3

卷积核的感受野为 7×7。 由此可见，空洞卷积可扩大卷积核的感受野，且不会增加卷积核的参数数量。

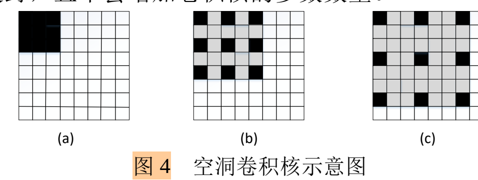


图 4　空洞卷积核示意图

Fig. 4　Schematic diagram of dilated convolutional filter

2.2　D-MobileNet 模型

感受野是指卷积神经网络每一层输出的特征图中每个元素在输入图像上映射的区域大小。层数越往后，感受野区域越大，越接近全局感受野。本文通过扩张局部感受野来提高 MobileNet 的分类精度，所以增加感受野的层数应靠近输入层。根据空洞卷积核所在位置的不同，提出了三种改进的网络模型，分别为 D1-MobileNet、D2-MobileNet 以及 D3-MobileNet。

a) D1-MobileNet。D1-MobileNe 将 MobileNet 的第一层卷积层步长设置为 1，并使用扩张率为 2 的空洞卷积核代替标准的卷积核。同时，为了增加最少的计算量，将第二层深度可分离卷积层中的深度卷积层步长设置为 2，其他层不变。这样，与 MobileNet 相比较，由于第一层的卷积步长设置为了 1，第一层卷积层输出的特征图大小由 112×112 变为 224×224，如图 5 所示。

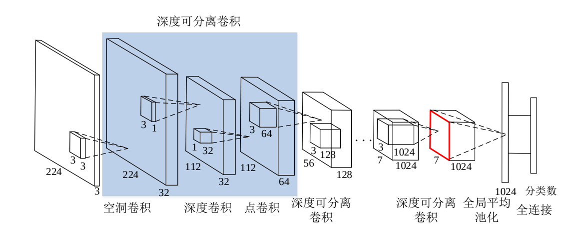


图 5　D1-MobileNet 网络结构图

Fig. 5　Architecture of D1-mobilenet

b) D2-MobileNet。在 MobileNet 的第二层深度可分离卷积层的深度卷积层中，用扩张率为 2 的空洞卷积核代替标准的卷积核，其他层不变。该方法不增加任何计算量和参数数量，也不改变任何一层的输出特征图大小，如图 6 所示。

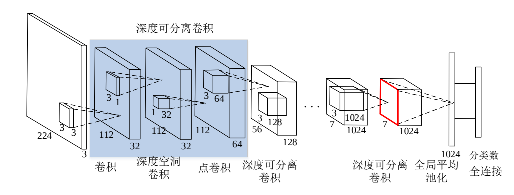


图 6　D2-MobileNet 网络结构图

Fig. 6　Architecture of D2-mobilenet

c) D3-MobileNet 将 MobileNet 的第一层卷积层步长设置为 1，用扩张率为 2 的空洞卷积核代替标准的卷积核， 并在第一层的批次规范化层[16](Batch Normalization)后加入步长为 2 的池化层，其他层不变，如图 7 所示。

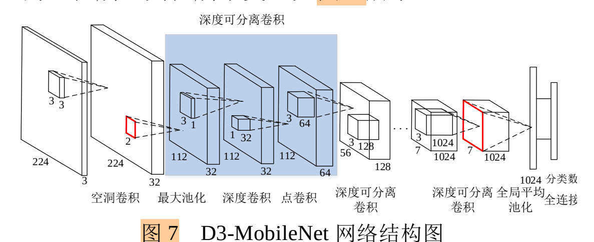


图 7　D3-MobileNet 网络结构图

Fig. 7　Architecture of D3-mobilenet

2.3　D-MobileNet 性能分析

对于一个标准的卷积层，假设输入 $R^{h\times w\times m}$ 的特征图 $\boldsymbol{I}$，其中 h、w、m 分别代表特征图的高、宽和输入特征图的通道数。$\boldsymbol{I}$ 与 $R^{s\times s\times m\times n}$（$\boldsymbol{s}$ 代表卷积核的尺寸，$\boldsymbol{n}$ 代表输出特征图的通道数）的卷积核 $\boldsymbol{K}$ 进行无填充的卷积操作，可得到 $R^{(h-s+1)\times(w-s+1)\times n}$ 的输出特征图 $\boldsymbol{O}$，$\boldsymbol{O}=\boldsymbol{K}*\boldsymbol{I}$。

$$O(y,x,j)=\sum_{i=1}^{m}\sum_{u,v=1}^{s}K(u,v,i,j)I(y+u-1,x+v-1,i) \tag{1}$$

其中：$O(y,x,j)$ 代表第 $\boldsymbol{j}$ 个特征图中点 (y,x) 的值，$K(u,v,i,j)$ 代表第 $\boldsymbol{j}$ 个卷积核中第 $\boldsymbol{i}$ 个通道上点 (u,v) 上的值， $I(y,x,i)$ 代表第 $\boldsymbol{i}$ 个输入通道上点 (y,x) 的值。由式(1)可知，得到一个输出需要 $s\times s\times m$ 次乘法运算，则总计算量为 $s\times s\times m\times(h-s+1)\times(w-s+1)\times n$，总的参数数量为 $s\times s\times m\times n$。

D-MobileNet 在标准卷积层引入空洞卷积核，输入同样的特征图 $\boldsymbol{I}$，用扩张率为 $\boldsymbol{r}$ 同尺寸大小的卷积核 $\boldsymbol{K}$ 进行的无填充的空洞卷积操作，可得到 $R^{(h-s-(s-1)(r-1)+1)\times(w-s-(s-1)(r-1)+1)\times n}$ 的输出特征图 $\boldsymbol{O_d}$。

$$O_d(y,x,j)=\sum_{i=1}^{m}\sum_{u,v=1}^{s}K(u,v,i,j) I(y+u+(u-1)(r-1)-1,x+v+(v-1)(r-1)-1,i) \tag{2}$$

由式(2)可知，空洞卷积层的总计算量为 $s\times s\times m\times(h-s-(s-1)(r-1)+1)\times(w-s-(s-1)(r-1)+1)\times n$， 参数数量为 $s\times s\times m\times n$。在无填充的卷积操作条件下，扩张率 $r>1$ 的空洞卷积的计算量要小于标准的卷积，参数数量一样，但空洞卷积的感受野比标准卷积大；在有填充的卷积操作条件下，输出特征图尺寸均为 $R^{h\times w\times n}$， 两者的计算量和参数数量都一样。

D-MobileNet 在深度可分离卷积层引入空洞卷积核， 输入特征图经过深度卷积层得到 $\boldsymbol{O_{dc}}$，再经过点卷积层， 最后可得到 $R^{(h-s-(s-1)(r-1)+1)\times(w-s-(s-1)(r-1)+1)\times n}$ 的输出特征图 $\boldsymbol{O}$。

$$O_{dc}(y,x,j)=\sum_{u,v=1}^{s}K(u,v,j) * I(y+u+(u-1)(r-1)-1,x+v+(v-1)(r-1)-1,j) \tag{3}$$

其中：$O_{dc}(y,x,j)$ 代表第 $\boldsymbol{j}$ 个特征图中点 (y,x) 的值，$K(u,v,j)$ 代表第 $\boldsymbol{j}$ 个卷积核上点 (u,v) 的值， $I(y,x,j)$ 代表第 $\boldsymbol{j}$ 个输入通道上点 (y,x) 的值。深度可分离卷积的总计算量为 $(s\times s+n)\times(h-s-(s-1)(r-1)+1)\times(w-s-(s-1)(r-1)+1)\times m$，总的参数数量为 $s\times s\times m+m\times n$。由此可见，深度可分离卷积层相对于标准卷积层，参数减少量为

$$\frac{s\times s\times m+m\times n}{s\times s\times m\times n}=\frac{1}{n}+\frac{1}{s^2} \tag{4}$$

计算量减少为

$$\frac{(s\times s+n)\times(h-s-(s-1)(r-1)+1)\times(w-s-(s-1)(r-1)+1)}{s\times s\times n\times(h-s+1)\times(w-s+1)} \tag{5}$$

同理，在进行有填充的深度可分离空洞卷积时，计算量减少为

$$\frac{(s\times s+n)\times m\times h\times w}{s\times s\times m\times n\times h\times w}=\frac{1}{n}+\frac{1}{s^2} \tag{6}$$

由式(2)和 (3)的输出特征图尺寸可知，扩张率为 $\boldsymbol{r}$ 卷积核大小为 $\boldsymbol{S\times S}$ 的深度卷积核 $\boldsymbol{K}$ 的感受野相当于卷积核 $(r\times s-r+1)\times(w\times s-r+1)$ 的感受野，可达到扩大感受野的目的，而且不会增加参数量和计算量

3　实验及结果分析

实验采用 TensorFlow 框架下的 Python 语言，模型在配有 NVIDIA TITAN GPU 的服务器上实现。实验采用 RMSprop

优化算法进行优化。 RMSprop 是一种自适应学习率方法，可调整学习率，初始学习率为 0.1。根据数据集训练样本数量的不同，本文设置不同的 epoch 数来降低学习率。权重初始化采用 Xavier 初始化方法，该方法可根据每层输入个数和输出个数来决定参数随机初始化分布范围， 是一种均匀分布，偏差初始值全为零。实验共训练 5 万批次， 每批样本数为 64，均采用 ReLU 作为激活函数。为了证明 D-MobileNet 模型的有效性，实验将 D-MobileNet 模型与 MobileNet 模型在 Caltech-101[22]、图宾根大学动物分类数据库和 Caltech-256[23]数据集上的分类结果进行比较。

3.1　Clatech-101 数据集

Caltech-101 数据集总共有 9 145 张图像，共 102 类。其中包含 101 个物体类别和一个背景类，每类图像的数量在 40~800 个，图 8 为 Caltech-101 数据集中的图片事例。在网络训练时，首先将数据集中的图片进行标签，然后充分打乱，随机选取其中的 1500 张图片作为测试集，剩余的图片作为训练集训练网络。



图 8　Caltech-101 数据集图片事例

Fig. 8　Picture instances in the Caltech-101 dataset

3.2　Clatech-256 数据集

Caltech-256 数据集在 Caltech-101 数据集的基础上增加了图像类别和每类图像的数量，总共 30607 张图像， 共 257 类。其中包含 256 个物体类别和一个背景类，每类图片最少 80 张，最多 827 张（背景类）， 图 9 为 Caltech-256 数据集中的图片事例。在训练网络时，将数据集中每张图片进行标签，然后打乱，随机抽取其中的 3060 张图片作为测试集， 剩余的图片作为训练集训练网络。

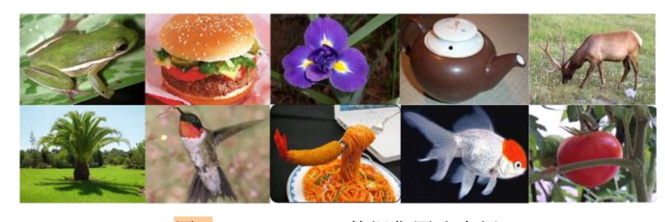


图 9　Caltech-256 数据集图片事例

Fig. 9　Picture instances in the Caltech-256 dataset

3.3　图宾根大学动物分类数据库

图宾根大学动物分类数据库(Uebingen animals with attributes)总共有 50 种动物类别，共 30 475 张图片。由于类别中的图片量差别大，实验选取其中最多的并且类别数目差别不大的 21 种动物类别作为数据集，共有 22 742 张图片，每类图片数量在 850~1600， 图 10 为图宾根大学动物数据集中的图片事例。在训练网络前，对数据集中的图片进行标注并随机抽取 2 000 张图片作为测试集，其余图片作为训练集训练网络。



图 10　Uebingen animals(21 类)数据集图片事例

Fig. 10　Picture instances in the Uebingen animals(21) dataset

3.4　实验结果分析

为了验证改进的有效性，将实验分成 4 组在相同运行环境和超参数数值的前提下进行结果分析和比较。第一组用标准的 MobileNet 神经网络结构进行图像分类，第二组用改进的 D1-Mobilenet 神经网络结构进行图像分类，第三组用改进的 D2-MobileNet 神经网络结构进行图像分类，第四组用改进的 D3-Mobilenet 神经网络结构进行图像分类。图 11 是四种分类方法在 Caltech-101 数据集上取得的分类正确率， 表 1 为相应的分类正确率数值。

表 1　Caltech-101 数据集上的准确率%

Table 1　Accuracy rate on Caltech-101101　dataset (%)

迭代次数	30000	35000	40000	45000	50000
MobileNet	76.73	76.6	76.6	76.8	76.6
D1_MobileNet	77.4	77.47	77.53	77.4	77.47
D2_MobileNet	77.67	77.8	77.73	77.67	77.73
D3_MobileNet	78.6	78.6	78.53	78.53	78.73

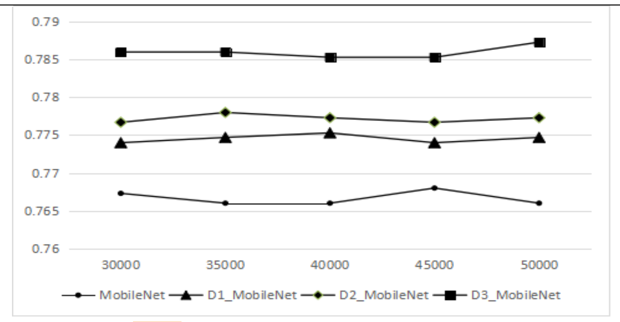


图 11　Caltech-101 数据集上的准确率

Fig. 11　Accuracy rate on Caltech-101 dataset

由图 11 和表 1 可知，四种分类模型在迭代 30000 次以后，其准确率均已达到平衡，且改进的三种模型其准确率均比 MobileNet 模型提高 0.8%~2%左右。其中，D1_MobileNet 模型可以提高 0.87%，D2_MobileNet 模型可以提高 1.13%，D3_MobileNet 模型精度提高最多，可以提高 2.13%，最终分类精度为 78.73%。

图 12 是四种分类方法在 Caltech-256 数据集上取得的分类正确率的比较，表 2 为相应的分类正确率数值。

表 2　Caltech-256 数据集上的准确率%

Table 2　Accuracy rate on Caltech-256　dataset (%)

迭代次数	30000	35000	40000	45000	50000
MobileNet	64.48	64.58	64.55	64.67	64.52
D1_MobileNet	65.77	65.74	65.87	65.9	65.87
D2_MobileNet	66.1	66.06	65.94	65.84	65.94
D3_MobileNet	64.97	64.9	64.87	65.19	65.16

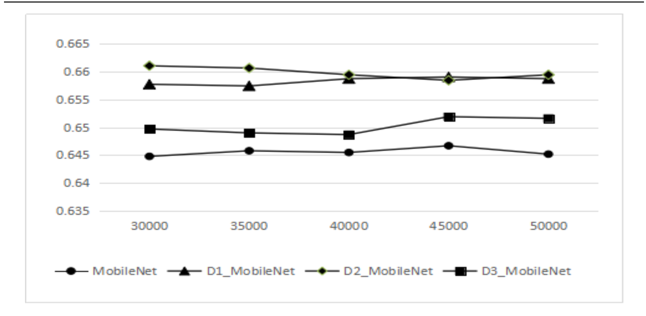


图 12　Caltech-256 数据集上的准确率

Fig. 12　Accuracy rate on Caltech-256 dataset

由图 12 和表 2 可知， 四种分类模型在迭代 30000 次以后， 其准确率均已达到平衡， 且改进的三种模型其准确率均比 MobileNet 模型提高 0.5%~1.5%左右。其中，D1_MobileNet 模型可以提高 1.35%，D3_MobileNet 模型可以

提高 0.64%， D2_MobileNet 模型精度提高最多，可以提高 1.42%，最终分类精度为 65.94%。

图 13 是五种分类方法在 Uebingen Animals 数据集上取得的分类正确率的比较，表 3 为相应的分类正确率数值。

表 3 Uebingen Animals(21 类)数据集上的准确率%

Table 3. Accuracy rate on Uebingen animals(21) dataset

迭代次数	30000	35000	40000	45000	50000
MobileNet	91.6	91.6	91.6	91.55	91.6
D1_MobileNet	92.45	92.45	92.5	92.35	92.4
D2_MobileNet	92.0	92.05	92.05	92.0	92.0
D3_MobileNet	92.85	92.75	92.8	92.7	92.8

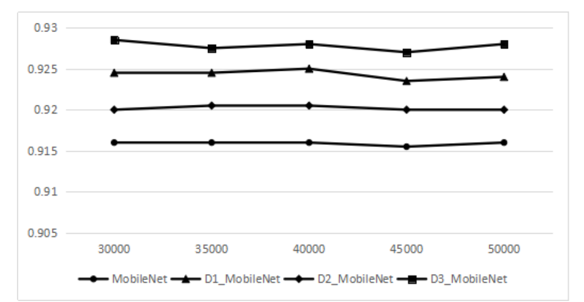


图 13 Uebingen Animals(21 类)数据集上的准确率

Fig. 13 Accuracy rate on Uebingen animals(21) dataset

由图 13 和表 3 可知，四种模型在迭代 30000 次时均达到平衡，准确率变化不大，且改进的 D-MobileNet 准确率比 MobileNet 高 0.5%~1.2%左右。其中 D1_MobileNet 模型最终提高 0.8%，D2_MobileNet 模型最终提高 0.4%，D3_MobileNet 模型的准确率提高最多，达到 1.2%，最终的分类精度为 92.8%。

4 结束语

深度学习的内存密集型和高度计算密集型特点使其在应用设备上的应用受到限制，而对网络模型进行压缩与加速，会损失分类精度。本文将空洞卷积与特殊的轻量级神经网络模型 MobileNet 结合，在不增加网络参数的前提下提高分类精度，使该轻量级网络更好的应用于低内存设备中。实验结果表明改进后的 D-MobileNet 在实验数据集上有更好的分类精度。

参考文献：

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] Wang Naiyan, Yeung D Y. Learning a deep compact image representation for visual tracking [C]// Proc of the 26th International Conference on Neural Information Processing Systems.[S.l.]: Curran Associates Inc,2013: 809-817.

[3] Wan J, Wang D, Hoi S C H, *et al.* Deep learning for content-based image retrieval: A comprehensive study [C]// Proc of the 22nd ACM International Conference on Multimedia.New York: ACM Press, 2014: 157-166.

[4] Chen L C, Papandreou G, Kokkinos I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2018, 40 (4): 834-848.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//Proc of IEEE International Conference on Computer Vision and Pattern Recognition. 2014.

[6] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[7] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep Residual Learning for Image Recognition [C]/ Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[8] Denil M, Shakibi B, Dinh L, *et al.* Predicting parameters in deep learning [C]// Proc of the 26th International Conference on Neural Information Processing Systems.[S.l.]: Curran Associates Inc, 2013: 2148-2156.

[9] Denton E, Zaremba W, Bruna J, *et al.* Exploiting linear structure within convolutional networks for efficient evaluation [C]// Proc of the 27th International Conference on Neural Information Processing Systems.Cambridge,MA:MIT Press, 2014: 1269-1277.

[10] Han Song, Mao Huizi, Dally W J. Deep compression: compressing deep neural networks with pruning, trained quantization and huffman coding [EB/OL].(2016-02-15). https://arxiv.org/pdf/1510.00149.pdf.

[11] Hinton G E, Vinyals O, Dean J. Distilling the knowledge in a neural network [EB/OL].(2015-03-09). https://arxiv.org/abs/1503.02531.

[12] Iandola F N, Han S, Moskewicz M W, *et al.* SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0. 5MB model size[EB/**OL]**. (2016-11-04). https://arxiv.org/pdf/1602.07360v3.pdf.

[13] Howard A G, Zhu Menglong, Chen Bo, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17). https://arxiv.org/abs/1704.04861.

[14] Zhang Xiangyu, Zhou Xinyu, Lin Mengxiao, *et al.* ShuffleNet: an extremely efficient convolutional neural network for mobile devices]// Proc of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 6848-6856.

[15] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. [EB/OL](2016-04-30). https://arxiv.org/abs/1511.07122.

[16] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// Proc of International Conference on International Conference on Machine Learning. 2015: 448-456.